

Mineração de Grafos e Predição de Links

Antonio Pecli

Ronaldo Goldschmidt

CONTEXTO

- Crescimento de aplicações de modelos de grafos para resolução de problemas do mundo real (biologia, economia, sistemas de recomendação, etc);
- Crescimento de diversas redes sociais (Orkut, Facebook, Twitter, Flickr, etc);
- Surgimento de diversas áreas de estudo: redes complexas (complex networks), análise de redes sociais (social network analysis), análise de links (link analysis);
- Necessidade de descobrir conhecimento a partir dos conjuntos de dados representados por estruturas e modelos de grafos.



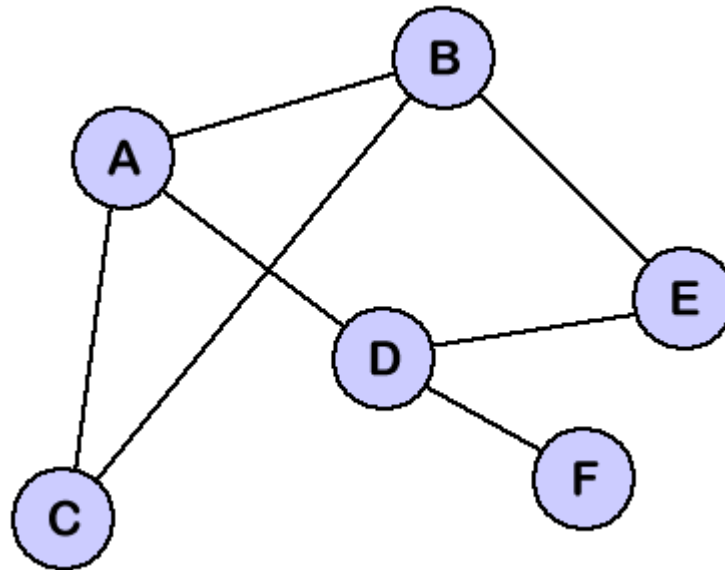
MINERAÇÃO DE GRAFOS

- **Mineração de grafos:** consiste na busca por padrões de relacionamento e comportamento ou novos conhecimentos em conjuntos de dados representados em modelos de grafos;
- Elementos representados em grafos geralmente possuem diversos relacionamentos (arestas) entre si, e estes podem possuir padrões importantes referente à maneira que esses elementos interagem no mundo real (ex: redes PPI – interação entre proteínas);
- Áreas como análise de redes complexas estudam a dinâmica e estrutura de diversos modelos de grafos em problemas do mundo real – onde existem diversas aplicações de técnicas de mineração de grafos;

MINERAÇÃO DE GRAFOS

Conceitos básicos: Grafos

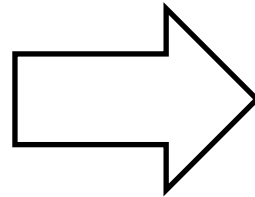
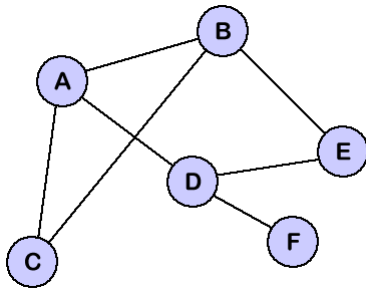
- Um grafo G é uma estrutura formada por dois conjuntos (V, A) , tais que:
 - V é o conjunto (não vazio) de **vértices** ou **nós** do grafo;
 - A é o conjunto de pares ordenados de **arestas** $a = (u, v)$, u e $v \in V$;



MINERAÇÃO DE GRAFOS

Representação dos grafos: Matriz e Lista de adjacência:

- Uma lista de adjacência é um conjunto que representa quais nós da rede possuem aresta com um determinado nó;
- Uma matriz de adjacência é uma matriz de tamanho $N \times N$, sendo N o número de nós do grafo, que representa quais as arestas existentes no grafo.



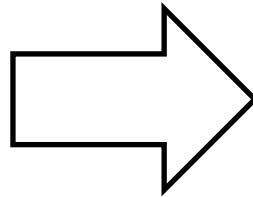
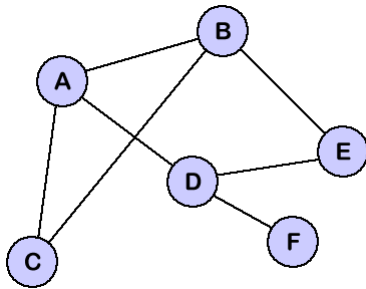
Nós	Lista de adjacência
A	{B, C, D}
B	{A, B, E}
C	{A, B}
D	{A, E, F}
E	{B, D}
F	{D}

Nós	A	B	C	D	E	F
A	0	1	1	1	0	0
B	1	0	1	0	1	0
C	1	1	0	0	0	0
D	1	0	0	0	1	1
E	0	1	0	1	0	1
F	0	0	0	1	0	0

MINERAÇÃO DE GRAFOS

Métricas (Centralidade)

- Centralidade de grau: fórmula
- Centralidade de proximidade: fórmula
- Centralidade de intermediação: fórmula



Nós	Centralidade de grau	Centralidade de proximidade	Centralidade de intermediação
A			
B			
C			
D			
E			
F			

MINERAÇÃO DE GRAFOS

Outros tipos de métricas

- Topológicas
 - Lista
- Caminhos
 - Lista
- Atributos dos nós
 - Exemplos

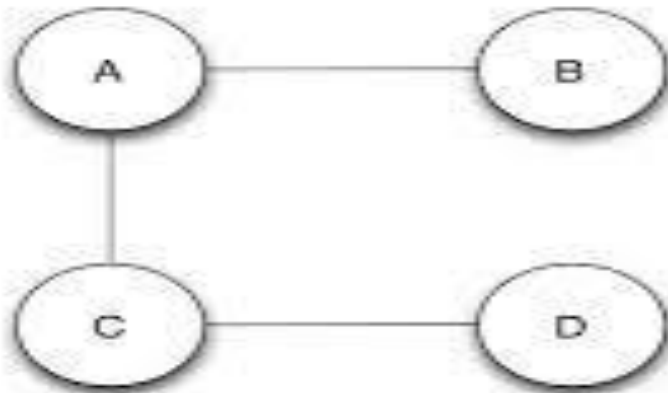
TAREFAS

Tarefas de mineração de grafos:

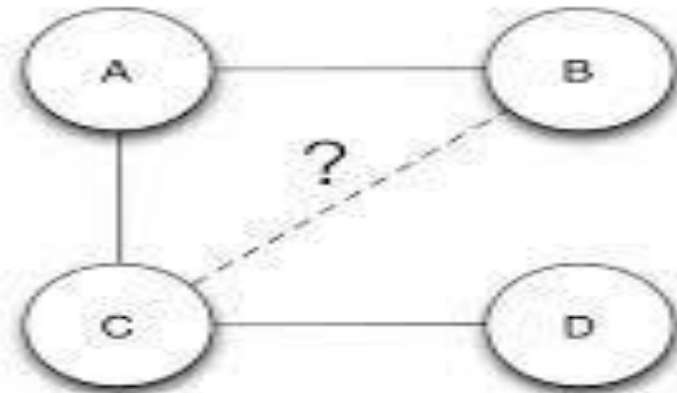
- Detecção de comunidades
- Rankeamento
- Predição de links

PREDIÇÃO DE LINKS

- Predição de links
- Técnica para tratamento de links ausentes, que consiste na aplicação de técnicas e métodos preditivos para prever links que não estão atualmente representados em uma rede.



Time T



Time T+1

FERRAMENTAS EXISTENTES

- Linguagem de programação: Python;
- SNAP (Stanford Network Analysis Platform);
- Scikit-learn (framework de aprendizado de máquina);

REFERÊNCIAS

- BILGIC, M., NAMATA, G. M., GETOOR, L. **Combining Collective Classification and Link Prediction.** Washington: ICDMW '07 Proceedings of the Seventh IEEE International Conference on Data Mining Workshops, pp. 381-386, 2007.
- BAGAI, R., SUNDERRAMAN, R. **A Paraconsistent Relational Data Model.** London: International Journal of Computer Mathematics, Gordon and Breach Science Publishers, vol. 55, no. 1, pp. 39-55, 1995.
- DAHABIAH, A., PUENTES, J., SOLAIMAN, B. **Digestive Casebase Mining Based on Possibility Theory and Linear Unidimensional Scaling.** Winsconsin: AIKED'09 Proceedings of the 8th WSEAS international conference on Artificial intelligence, knowledge engineering and data base, 2009.
- DOPPA, J., YU, J., TADEPALLI, P., GETOOR, L. **Chance-Constrained Programs for Link Prediction.** NIPS Workshop on Analyzing Networks and Learning with Graphs, 2009.